Repulsive-SVDD Classification

Phuoc Nguyen and Dat Tran(✉)

Faculty of Education, Science, Technology and Mathematics,
University of Canberra, Canberra, ACT 2601, Australia
dat.tran@canberra.edu.au

Abstract. Support vector data description (SVDD) is a well-known kernel method that constructs a minimal hypersphere regarded as a data description for a given data set. However SVDD does not take into account any statistical distribution of the data set in constructing that optimal hypersphere, and SVDD is applied to solving one-class classification problems only. This paper proposes a new approach to SVDD to address those limitations. We formulate an optimisation problem for binary classification in which we construct two hyperspheres, one enclosing positive samples and the other enclosing negative samples, and during the optimisation process we move the two hyperspheres apart to maximise the margin between them while the data samples of each class are still inside their own hyperspheres. Experimental results show good performance for the proposed method.

Keywords: Repulsive SVDD · Support vector data description · Support vector machine · Classification

1 Introduction

Support vector data description (SVDD) [1] was proposed by Tax and Duin to train a hyperspherically shaped boundary around a normal dataset while keeping all abnormal data samples outside the hypersphere. This SVDD has been a successful approach to solving one-class problems such as outlier detection since the volume of this data description is kept minimal. One-class support vector machine (OC-SVM) [2] is a similar approach proposed earlier to estimate the support of a high-dimensional distribution. Although this method uses a maximal-margin hyperplane instead of a hypersphere to separate the normal data from the abnormal data, it has the same optimisation problem as SVDD. In both OC-SVM and SVDD, the boundary in the feature space when mapped back to the input space can produce a complex and tight description of the data distribution.

There are various extensions to SVDD. A small hypersphere and large margin approach was proposed in [3] for novelty detection problems where a minimal hypersphere was trained to include most of normal examples while the margin between the hypersphere and outliers is as large as possible. A further extension using two large margins instead of one was proposed in [4], where an interior margin between the hypersphere and the normal data and an exterior margin

T. Cao et al. (Eds.): PAKDD 2015, Part I, LNAI 9077, pp. 277–288, 2015.
DOI: 10.1007/978-3-319-18038-0_22

between the hypersphere and the abnormal data both are maximised. In [5], the authors define an optimisation problem as maximising the separation ratio $(R+d)/(R-d)$, where R is the hypersphere's radius and d is the hypersphere's margin. It is shown to be equivalent to minimising $(R^2 - kd^2)$ where k is a parameter to adjust between minimising R and maximising d. Hao et al. [6] also used a similar formulation in which several similarity functions were used to compute the distance to centres. Another extension of SVDD is [7] in which the use of two SVDDs for the description of data with two classes was proposed.

However all of those models are for one-class problems in which the task is to provide a tight data description or to detect outliers. When applying to a two-class problem where the numbers of data samples of two classes are not much different, the boundary of one-class methods is inappropriate. To overcome this problem, the first straight forward approach is to train two SVDDs, one for each class and define the decision boundary as the bisector between two surfaces of the hyperspheres. Although this approach improves the performance of one-class methods for two-class problems, they are limited by the small-sphere constraint of the data description.

In this paper, we propose a method using two SVDDs, one enclosing positive samples and the other enclosing negative samples, for binary classification tasks. The minimum bounding hypersphere constraint is relaxed to allow the hyperspheres to acquire larger regions. This is achieved by imposing a criterion that maximises the distance between two hyperspheres while still keeping the data inside the spheres. A margin variable is added to the optimisation to further improve the classification boundary. Since the proposed method trains two SVDDs that repel each other, we call it repulsive-SVDD classification (RSVC). RSVC decision boundary can be considered as a compromise between the boundary of a SVM boundary and a bisector boundary of two SVDDs' surfaces, this is controlled by a trade off parameter to adjust the balance between describing the data and maximising the distance between the two sphere centres.

The rest of the paper is organized as follows. The theory of the proposed RSVC will be presented in Section 2. Comparison of RSVC with Two SVDDs will be discussed in Section 3. Experimental results are presented to show the performance of the proposed method in Section 4. Finally, Section 5 presents our conclusions.

2 Proposed Approach: Repulsive-SVDD Classification (RSVC)

To apply SVDD for binary classification problems, we construct a hypersphere for each class to describe its data distribution with additional properties to discriminate the two classes. First, the hypersphere constraint in SVDD is relaxed to allow this hypersphere to acquire a larger area that is far from the other class. This is achieved by imposing a criterion that maximises the distance between two hyperspheres while still keeping all data samples of a class inside its hypersphere. Second, the margin (i.e., the distance between surfaces of the two hypersphere) is maximised, similar to the maximal margin philosophy of a support vector machine.

A visualisation of RSVC is demonstrated in Fig. 1. In the left figure, SVM determines a maximum margin hyperplane without considering data distributions of positive and negative classes. Whereas in the middle figure, SVDDs determine two minimal hyperspheres without considering the margin between the two classes, and the decision boundary is the perpendicular hyperplane of the line segment connecting the two hypersphere centres.

By contrast, our RSVC can provide an intermediate solution between SVM and SVDDs. Given the problem in Fig. 1, the RSVC optimisation problem attempts to keep the radii minimum while maximising the distance between the two hyperspheres. As a result, the hyperspheres will expand in the direction that increases the distance between the two hyperspheres. Moreover, the weights of these two directions can be controlled by a parameter.

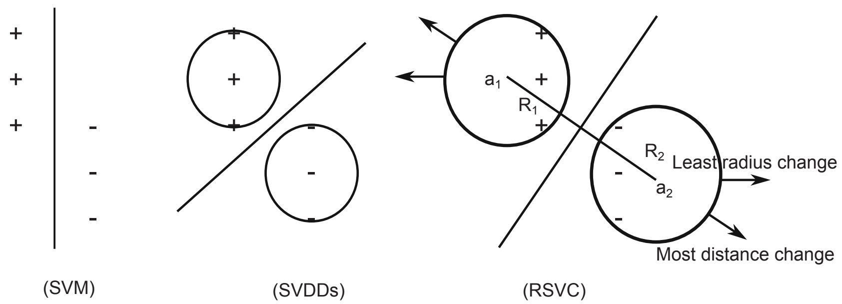

Fig. 1. SVM (left figure) determines a maximum margin hyperplane without considering data distributions of positive and negative classes. SVDDs (middle figure) determine minimum hyperspheres without considering the margin between two classes. RSVC (right figure) determines two minimal hyperspheres, one enclosing positive samples and the other enclosing negative samples, while maximising the distance between two centres to a degree controlled by a parameter.

2.1 Problem Formulation

Consider a dataset $\{x_i\}, i = 1, \ldots, n$ with two classes, positive class with n_1 data samples and negative class with n_2 data samples, $n_1 + n_2 = n$. The problem of RSVC is to determine two optimal hyperspheres (a_1, R_1) and (a_2, R_2), one encloses data samples of the positive class and the other encloses data samples of the negative class, and at the same time maximise the distance between the two centres. In addition, all positive and negative data samples are forced to stay outside the margin ρ_1 and ρ_2 of the positive hypersphere and the margin of the negative hypersphere respectively. The optimisation problem is formulated as follows:

$$\min_{R_1,R_2,a_1,a_2,\rho_1,\rho_2} \quad R_1^2 + R_2^2 - k||a_1 - a_2||^2 - \mu_1\rho_1 - \mu_2\rho_2 \tag{1}$$

$$s.t. \quad ||\phi(x_i) - a_1||^2 \leq R_1^2 - \rho_1, \qquad \forall i, y_i = +1 \tag{2}$$

$$||\phi(x_i) - a_1||^2 \geq R_1^2 + \rho_1, \qquad \forall i, y_i = -1 \tag{3}$$

$$||\phi(x_i) - a_2||^2 \leq R_2^2 - \rho_2, \qquad \forall i, y_i = -1 \tag{4}$$

$$||\phi(x_i) - a_2||^2 \geq R_2^2 + \rho_2, \qquad \forall i, y_i = +1 \tag{5}$$

$$\rho_1 \geq 0, \rho_2 \geq 0 \tag{6}$$

where k is a parameter which represents the repulsive degree between two centres, μ_1 and μ_2 are two parameters controlling the support vectors, and ϕ is the mapping to transform the vector x_i to a feature space.

The above problem is for separable datasets. In practice, to allow errors, the constraints are relaxed by introducing slack variables ξ_{1i} and ξ_{2i}, and penalized terms are added to its objective function. In addition, if we combine the constraints in this problem to have a simpler form, the optimisation problem becomes:

$$\min_{R_1,R_2,a_1,a_2,\rho_1,\rho_2,\xi_{1i},\xi_{2i}} \quad R_1^2 + R_2^2 - k||a_1 - a_2||^2 - \mu_1\rho_1 - \mu_2\rho_2 + \frac{1}{\nu_1 n_1}\sum_i \xi_{1i} + \frac{1}{\nu_2 n_2}\sum_i \xi_{2i} \tag{7}$$

$$s.t. \quad y_i||\phi(x_i) - a_1||^2 \leq y_i R_1^2 - \rho_1 + \xi_{1i}, \qquad \forall i \tag{8}$$

$$y_i||\phi(x_i) - a_2||^2 \geq y_i R_2^2 + \rho_2 - \xi_{2i}, \qquad \forall i \tag{9}$$

$$\rho_1 \geq 0, \rho_2 \geq 0 \tag{10}$$

$$\xi_{1i} \geq 0, \xi_{2i} \geq 0 \qquad \forall i \tag{11}$$

where ν_1 and ν_2 are parameters controlling the number of support vectors, together with μ_1 and μ_2. They will be explained in Proposition 1 below.

2.2 Convex Formulation of RSVC

Although the optimisation in (7) has a non-convex objective function, it can be reformulated to have a convex form as follows:

$$\min_{R_1,R_2,a_1,a_2,\rho_1,\rho_2,\xi_{1i},\xi_{2i}} \quad R_1^2 - a_1^2 + R_2^2 - a_2^2 + a_1^2 + a_2^2 - k||a_1 - a_2||^2 - \mu_1\rho_1 - \mu_2\rho_2 + \frac{1}{\nu_1 n_1}\sum_i \xi_{1i} + \frac{1}{\nu_2 n_2}\sum_i \xi_{2i} \tag{12}$$

$$s.t. \quad y_i\phi(x_i)^2 - 2y_i\phi(x_i)a_1 \leq y_i(R_1^2 - a_1^2) - \rho_1 + \xi_{1i}, \forall i \tag{13}$$

$$y_i\phi(x_i)^2 - 2y_i\phi(x_i)a_2 \geq y_i(R_2^2 - a_2^2) + \rho_2 - \xi_{2i}, \forall i \tag{14}$$

$$\rho_1 \geq 0, \rho_2 \geq 0 \tag{15}$$

$$\xi_{1i} \geq 0, \xi_{2i} \geq 0 \qquad \forall i \tag{16}$$

Let $\delta_1 = a_1^2 - R_1^2, \delta_2 = a_2^2 - R_2^2$ and $0 \le \delta_0 \le ||a_1 - a_2||^2$, (12) becomes

$$\begin{aligned}
\min_{\delta_1,\delta_2,\delta_0,a_1,a_2,\rho_1,\rho_2,\xi_{1i},\xi_{2i}} \quad & -\delta_1 - \delta_2 + a_1^2 + a_2^2 - k\delta_0 - \mu_1\rho_1 - \mu_2\rho_2 \\
& + \frac{1}{\nu_1 n_1}\sum_i \xi_{1i} + \frac{1}{\nu_2 n_2}\sum_i \xi_{2i} && (17)\\
s.t. \quad & 2y_i\phi(x_i)a_1 - y_i\phi(x_i)^2 \ge y_i\delta_1 + \rho_1 - \xi_{1i}, \qquad \forall i && (18)\\
& 2y_i\phi(x_i)a_2 - y_i\phi(x_i)^2 \le y_i\delta_2 - \rho_2 + \xi_{2i}, \qquad \forall i && (19)\\
& \rho_1 \ge 0, \rho_2 \ge 0 && (20)\\
& \xi_{1i} \ge 0, \xi_{2i} \ge 0 \qquad \forall i && (21)\\
& ||a_1 - a_2||^2 \ge \delta_0 && (22)\\
& \delta_0 \ge 0 && (23)
\end{aligned}$$

We can construct the Lagrange function below using these following Lagrange multipliers $\alpha_{1i}, \alpha_{2i}, \gamma_{1i}, \gamma_{2i}, \theta_1, \theta_2, \beta, \lambda$:

$$\begin{aligned}
L(\delta_1,\delta_2,\delta_0,a_1,a_2,\rho_1,\rho_2,\xi_{1i},\xi_{2i},\alpha_{1i},\alpha_{2i},\gamma_{1i},\gamma_{2i},\theta_1,\theta_2,\beta,\lambda) = -\delta_1 - \delta_2 \\
+a_1^2 + a_2^2 - k\delta_0 - \mu_1\rho_1 - \mu_2\rho_2 + \frac{1}{\nu_1 n_1}\sum_i \xi_{1i} + \frac{1}{\nu_2 n_2}\sum_i \xi_{2i} \\
-\sum_i \alpha_{1i}(2y_i\phi(x_i)a_1 - y_i\phi(x_i)^2 - y_i\delta_1 - \rho_1 + \xi_{1i}) - \sum_i \gamma_{1i}\xi_{1i} - \theta_1\rho_1 \qquad (24)\\
+\sum_i \alpha_{2i}(2y_i\phi(x_i)a_2 - y_i\phi(x_i)^2 - y_i\delta_2 + \rho_2 - \xi_{2i}) - \sum_i \gamma_{2i}\xi_{2i} - \theta_2\rho_2 \\
-\beta(||a_1 - a_2||^2 - \delta_0) - \lambda\delta_0
\end{aligned}$$

Using KKT conditions, we have:

$$\frac{\partial L}{\partial \delta_1} = 0 \Rightarrow -1 + \sum_i \alpha_{1i}y_i = 0 \Rightarrow \sum_i \alpha_{1i}y_i = 1 \qquad (25)$$

$$\frac{\partial L}{\partial \delta_2} = 0 \Rightarrow -1 + \sum_i \alpha_{2i}y_i = 0 \Rightarrow \sum_i \alpha_{2i}y_i = 1 \qquad (26)$$

$$\frac{\partial L}{\partial \delta_0} = 0 \Rightarrow -k + \beta - \lambda = 0 \Rightarrow \beta - \lambda = k \qquad (27)$$

$$\frac{\partial L}{\partial a_1} = 0 \Rightarrow (1-\beta)a_1 + \beta a_2 = \sum_i \alpha_{1i}y_i\phi(x_i) = A \qquad (28)$$

$$\frac{\partial L}{\partial a_2} = 0 \Rightarrow (1-\beta)a_2 + \beta a_1 = -\sum_i \alpha_{2i}y_i\phi(x_i) = -B \qquad (29)$$

$$\frac{\partial L}{\partial \xi_{1i}} = 0 \Rightarrow \frac{1}{\nu_1 n_1} - \alpha_{1i} - \gamma_{1i} = 0 \Rightarrow \alpha_{1i} + \gamma_{1i} = \frac{1}{\nu_1 n_1} \quad \forall i \qquad (30)$$

$$\frac{\partial L}{\partial \xi_{2i}} = 0 \Rightarrow \frac{1}{\nu_2 n_2} - \alpha_{2i} - \gamma_{2i} = 0 \Rightarrow \alpha_{2i} + \gamma_{2i} = \frac{1}{\nu_2 n_2} \quad \forall i \qquad (31)$$

$$\frac{\partial L}{\partial \rho_1} = 0 \Rightarrow -\mu_1 + \sum_i \alpha_{1i} - \theta_1 = 0 \Rightarrow \sum_i \alpha_{1i} - \theta_1 = \mu_1 \tag{32}$$

$$\frac{\partial L}{\partial \rho_1} = 0 \Rightarrow -\mu_2 + \sum_i \alpha_{2i} - \theta_2 = 0 \Rightarrow \sum_i \alpha_{2i} - \theta_2 = \mu_2 \tag{33}$$

Equations (28) and (29) leads to

$$\begin{cases} a_1 + a_2 = A - B \\ a_1 - a_2 = \frac{A+B}{1-2\beta} \end{cases} \Rightarrow \begin{cases} a_1 = \frac{(1-\beta)A+\beta B}{1-2\beta} \\ a_2 = \frac{-\beta A+(\beta-1)B}{1-2\beta} \end{cases} \tag{34}$$

By substituting the KKT conditions into the Lagrangian function we obtain the dual form of the optimisation:

$$\min \frac{1}{1-2k}\Big[(1-k)\sum_{i,j} \alpha_{1i}\alpha_{1j}y_iy_jK(x_i,x_j) + (1-k)\sum_{i,j} \alpha_{2i}\alpha_{2j}y_iy_jK(x_i,x_j)$$
$$+2k\sum_{i,j} \alpha_{1i}\alpha_{2j}y_iy_jK(x_i,x_j)\Big] + \sum_i \alpha_{1i}y_iK(x_i,x_i) - \sum_i \alpha_{2i}y_iK(x_i,x_i) \tag{35}$$

$$s.t. \quad \sum_i \alpha_{1i}y_i = 1\ ,\ \sum_i \alpha_{2i}y_i = -1 \tag{36}$$

$$\sum_i \alpha_{1i} = \mu_1\ ,\ \sum_i \alpha_{2i} = \mu_2 \tag{37}$$

$$0 \le \alpha_{1i} \le \frac{1}{\nu_1 n_1}\ ,\ 0 \le \alpha_{2i} \le \frac{1}{\nu_2 n_2} \quad \forall i \tag{38}$$

where the inner product between vectors has been replaced by the kernel K, and the Lagrange multipliers $\gamma_{1i} \ge 0, \gamma_{2i} \ge 0, \theta_1 \ge 0, \theta_2 \ge 0, \lambda \ge 0$ have been removed from Equations (30), (31), (32), (33) and (27) respectively. Similarly to ν-SVC, $\sum_i \alpha_{1i}$ is set to μ_1, $\sum_i \alpha_{2i}$ is set to μ_2 and β is set to k, where k is a parameter chosen in the range $k \in [0, \frac{1}{2})$.

It can be seen that if k is set to 0 in the above optimisation problem then the RSVC optimisation problem (35) can be broken into two independent optimisation problems similar to SVDDs except for the extra constraints $\sum_i \alpha_{1i} = \mu_1$ and $\sum_i \alpha_{2i} = \mu_2$ resulting from the margin requirements in the original RSVC problem (1).

Solving the problem (35) gives a set of α_{1i}, α_{2i}. Then the centres a_1, a_2 can be determined from Equations (34).

To determine the radius R_1, the support vector x_t that lies on the surface of the hypersphere (a_1, R_1) and corresponds to the smallest $\alpha_{1t} \in (0, \frac{1}{\nu_1 n_1})$ is selected. Then the radius R_1 is calculated as $R_1 = d_1(x_t)$, where $d_1(x_t)$ is the distance from x_t to the centre a_1 and is determined as follows:

$$\begin{aligned} d_1^2(x_t) = \|\phi(x_t) - a_1\|^2 =& K(x_t,x_t) - \frac{2}{1-k}\big[(1-k)\sum_i \alpha_{1i}y_iK(x_t,x_i) \\ &+ k\sum_i \alpha_{2i}y_iK(x_t,x_i)\big] + a_1^2 \end{aligned} \tag{39}$$

The radius R_2 is calculated similarly:

$$d_2^2(x_t) = \|\phi(x_t) - a_2\|^2 = K(x_t, x_t) - \frac{2}{1-k}\Big[- k \sum_i \alpha_{2i} y_i K(x_t, x_i) \\ + (k-1) \sum_i \alpha_{2i} y_i K(x_t, x_i)\Big] + a_2^2 \tag{40}$$

In the test phase, a sample x can be determined whether it belongs to the hypersphere (a_1, R_1) or (a_2, R_2), i.e. class +1 or class -1, by the following decision function:

$$sign(d_2^2(x) - d_1^2(x)) \tag{41}$$

2.3 ν-Property

Following [8], a data sample x_i is called a *support vector* if it has Lagrange multiplier $\alpha_i > 0$; a data sample is called a *margin error* if it has positive slack variable $\xi_i > 0$.

Similarly to the property of the ν parameter in ν-SVC [8], we derive the property for the ν_1, ν_2, μ_1 and μ_2 parameters and use it for parameter selection to train the RSVC.

Proposition 1. *Let m_1 and m_2 denote the number of margin errors of the positive sphere and negative sphere respectively, and let s_1 and s_2 denote their numbers of support vectors. Then for parameters ν_1, ν_2, μ_1 and μ_2 we have:*

1. *$\mu_1\nu_1$ and $\mu_2\nu_2$ are upper bounds on the fraction of margin errors, and a lower bound on the fraction of support vectors for the positive sphere and negative sphere respectively:*

$$\frac{m_1}{n_1} \leq \mu_1\nu_1 \leq \frac{s_1}{n_1} \quad and \quad \frac{m_2}{n_2} \leq \mu_2\nu_2 \leq \frac{s_2}{n_2} \tag{42}$$

2. *The feasible ranges of ν_1, ν_2, μ_1 and μ_2 are:*

$$0 < \nu_1 \leq 1 \ , \ 1 \leq \mu_1 \leq \frac{1}{\nu_1} \quad and \quad 0 < \nu_2 \leq 1 \ , \ 1 \leq \mu_2 \leq \frac{1}{\nu_2} \tag{43}$$

Proof. We first prove for the positive hypersphere.

1. By the KKT conditions, all data points with $\xi_{1i} > 0$ imply $\gamma_{1i} = 0$. From (30) we have the equation $\alpha_{1i} = 1/(\nu_1 n_1)$ holds for every margin error. Summing up α_{1i} and using $\sum_i \alpha_{1i} = \mu_1$ from (37) we have:

$$\frac{m_1}{\nu_1 n_1} \leq \sum_i \alpha_{1i} = \mu_1 \tag{44}$$

On the other hand, (38) indicates that each support vector of the positive hypersphere can get at most $1/(\nu_1 n_1)$. Therefore summing up α_{1i} for support vectors of positive hypersphere, plus $\alpha_{1i} = 0$ for non-support vectors, and from (37) we have:

$$\frac{s_1}{\nu n_1} \geq \sum_i \alpha_{1i} = \mu_1 \tag{45}$$

Combining (44) and (45) we have the inequalities (42) for the positive hypersphere.

2. From (42) we have $0 < \mu_1 \nu_1 \leq 1$. In addition, from (36) we have $\sum_i \alpha_{1i} y_i = 1$, or $\sum_{\{i:y_i=+1\}} \alpha_{1i} = 1 + \sum_{\{i:y_i=-1\}} \alpha_{1i}$.
Since $\alpha_{1i} \geq 0 \ \forall i$, this leads to $\mu_1 = \sum_i \alpha_{1i} \geq \sum_{\{i:y_i=+1\}} \alpha_{1i} \geq 1$.
Combining these results we have the proof of (43).

The proof of inequalities (42) and (43) for the negative hypersphere is similar.

The proposed RSVC is for binary classification problems. It can be extended for multi-class classification problems by using "one-against-the rest" approach or "one-against-one" approach. Following [9], we use the one-against-one approach in this paper where data of every pair of classes are used to train a binary classifier that separates the two classes, resulting in $M(M-1)/2$ classifiers in a M-class classification problem. In the test phase, a voting strategy is used: each binary classification of a test sample generates a vote, and the class with the maximum number of votes for this test data sample is output as the overall classification result. In case that two classes have identical votes, one can simply choose the class appearing first in the array of storing class names as in [9].

3 Comparison of RSVC with Two SVDDs

SVDD can be extended to two SVDDs to describe a data set of two classes. Consider a data set $\{x_i\}, i = 1, \ldots, n$ of two classes, positive class with n_1 data samples and negative class with n_2 data samples, $n_1 + n_2 = n$. The optimisation problem is formulated as follows [7]:

$$\begin{aligned}
\min_{R_1,R_2,a_1,a_2,\xi_{1i},\xi_{2i}} \quad & R_1^2 + R_2^2 + \frac{1}{\nu_1 n_1}\sum_i \xi_{1i} + \frac{1}{\nu_2 n_2}\sum_i \xi_{2i} \\
s.t. \quad & ||x_i - a_1||^2 \leq R_1^2 + \xi_{1i}, \qquad \forall i, y_i = +1 \\
& ||x_i - a_1||^2 \geq R_1^2 - \xi_{1i}, \qquad \forall i, y_i = -1 \\
& ||x_i - a_2||^2 \leq R_2^2 + \xi_{2i}, \qquad \forall i, y_i = -1 \\
& ||x_i - a_2||^2 \geq R_2^2 - \xi_{2i}, \qquad \forall i, y_i = +1 \\
& \xi_{1i} \geq 0, \xi_{2i} \geq 0 \qquad \forall i
\end{aligned} \tag{46}$$

where (a_1, R_1) and (a_2, R_2) are two hyperspheres, ν_1, ν_2 are parameters.

This optimisation can produce a description of two minimal hyperspheres enclosing two classes. The decision boundary can be defined as the bisector between their surfaces. However this model is for one-class problems in which the task is to provide a tight data description or to detect outliers. When applying to a two-class problem where the data samples of two classes are balance the boundary of one-class methods is inappropriate. The RSVC can overcome this problem by allowing hyperspheres to acquire a larger area by minimising $-k||a_1-a_2||^2$ and creating a larger margin by minimising $-\mu_1\rho_1 - \mu_2\rho_2$ while still trying to provide data description for two classes.

4 Experiments

4.1 2-D Demonstration of RSVC

Figure 2 shows visual results for experiments performed on a simple 2-D datasets using RSVC. When parameter $k = 0$, the RSVC optimisation function becomes the optimisation function for two SVDDs, hence two SVDDs is a special case of RSVC. It can be seen that when k increases, two hyperspheres repulsed each other, resulting in a larger margin in between. Those data samples outside the hyperspheres but inside this margin are penalised by a cost proportional to $1/(\nu_1 n_1)$ or $1/(\nu_2 n_2)$. The decision boundary is the bisector between the hyperspheres' surfaces. The first row in Figure 2 shows that when parameter k increases, the hypersphere enclosing positive samples is moving away from negative samples while keeping all the positive samples inside it. The second row in Figure 2 shows that when $\mu_1\nu_1$ and $\mu_2\nu_2$ increase, more positive samples are outside the hyperspheres.

Classification experiments were conducted on 9 UCI datasets[1]. Details of these datasets are listed in Table 1. The datasets were divided in to 2 subsets, the subset contained 50% of the data is for training and the other 50% for testing. The training process was done using 5-fold cross validation. The parameters for the methods are as follows. Gaussian mixture models (GMM) [10] use 64 mixture components. OC-SVM parameters are searched in $\gamma \in \{2^{-13}, 2^{-11}, \ldots, 2^1\}$ and $\nu \in \{2^{-5}, 2^{-4}, \ldots, 2^{-2}\}$. Parameters of SVDD and SVDD with negative examples (Two SVDDs) are searched in $\gamma \in \{2^{-13}, 2^{-11}, \ldots, 2^1\}$ and $\nu \in \{2^{-5}, 2^{-4}, \ldots, 2^{-2}\}$. SVM parameters are search in $\gamma \in \{2^{-13}, 2^{-11}, \ldots, 2^1\}$ and $C \in \{2^{-1}, 2^3, \ldots, 2^{15}\}$; and RSVC parameters are searched in $\gamma \in \{2^{-7}, 2^{-5}, \ldots, 2^{-1}\}$, $\nu_1 = \nu_2 \in \{0.001, 0.01\}$, $\mu_1 = \mu_2 \in \{10, 30, \ldots, 90\}$, and $k \in \{0.5, 0.7, 0.9\}$.

Note that the parameter γ in RSVC is searched in a narrower range than that in SVM, while $\nu_1 n_1$ and $\nu_2 n_2$ are searched in a roughly similar number of options as of parameter C. This is to produce a sparse number of support vectors and avoid over fitting of the two SVDDs. Parameter $k \in \{0.5, 0.7, 0.9\}$ is to favour classification more than tight description. After the best parameters are selected in the cross validation step, the models are trained again with them on the whole training set and are tested on the 50% unseen test set. Experiments

[1] Available online at http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/

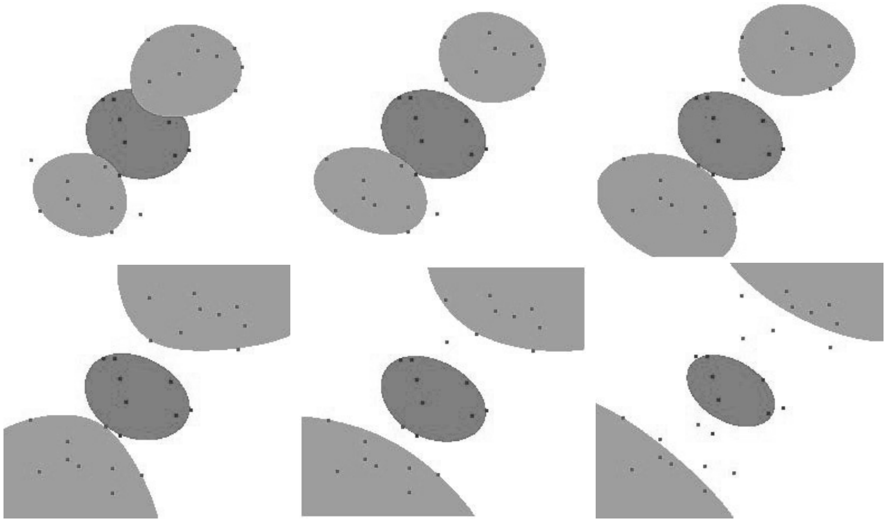

Fig. 2. The first row contains screenshots for RSVC when $k = 0, 0.3$ and 0.6, and $\mu_1\nu_1 = \mu_2\nu_2 = 0.2$. The second row contains screenshots for RSVC when $\mu_1\nu_1 = \mu_2\nu_2 = 0.1, 0.2$ and 0.5, and $k = 0.9$. A Gaussian RBF kernel was used, with $\gamma = 5$. Red points are positive samples and blue points are negative samples.

were repeated 10 times and the results were averaged with standard deviations given.

Table 2 shows the prediction rates in cross validation training. Table 3 shows the prediction rates on unseen test sets with best parameters selected.

It can be seen that the GMM, OCSVM and SVDD have undesirable performance in the classification task.

The two SVDDs have much higher performance than these one-class methods since they describe two minimal hyperspheres enclosing two classes and the

Table 1. Dataset information: number of classes, dataset size and number of features

Data set	#class	size	#feature
Fourclass	2	862	2
Liver disorders	2	345	6
Heart	2	270	13
Wine	3	178	13
Breast Cancer	2	683	10
Diabetes	2	768	8
Australian	2	690	14
Ionosphere	2	351	34
German numer	2	1000	24

decision boundary is the bisector between their surfaces. It can be seen that SVM has higher performance than two SVDDs, it trains a maximal-margin separating hyperplane rather than two minimal hyperspheres. RSVC show highest performance in most datasets. RSVC can overcome the limitation of two SVDDs for the classification task by training two SVDDs that repel each other, allowing spheres to acquire a larger area and creating a larger margin while still trying to provide data description for two classes.

Table 2. Prediction rates in cross validation training of classification methods

Dataset	GMM	OCSVM	SVDD	Two SVDDs	SVM	RSVC
Fourclass	67.24 ±5.73	62.15 ±3.3	54.01 ±4.97	71.44 ±4.15	75.08 ±4.4	77.98 ±4.52
Liver disorders	40.86 ±5.63	50.41 ±5.69	55.41 ±5.87	55.15 ±5.49	59.90 ±3.53	60.14 ±5.56
Heart	46.33 ±4.26	60.24 ±5.89	46.41 ±4.24	61.11 ±5.98	72.56 ±4.24	76.44 ±4.45
Wine	33.43 ±5.03	55.57 ±4.07	46.43 ±5.8	59.89 ±3.7	75.24 ±5.56	83.15 ±4.59
Breast cancer	56.52 ±3.34	73.85 ±4.11	62.91 ±4.15	77.16 ±4.08	81.29 ±4.44	81.49 ±4.46
Diabetes	55.24 ±5.06	51.84 ±3.99	40.24 ±5.16	50.95 ±5.63	63.47 ±5.93	66.87 ±3.28
Australian	54.15 ±5.84	58.36 ±5.52	48.23 ±5.16	61.03 ±5.75	70.96 ±3.53	71.90 ±3.66
Inosphere	57.12 ±3.61	65.48 ±3.85	34.26 ±3.48	68.92 ±4.59	73.69 ±4.51	75.86 ±4.29
German numer	40.09 ±5.49	58.96 ±5.39	58.14 ±5.31	59.65 ±5.51	64.04 ±5.99	65.75 ±3.35

Table 3. Prediction rates on unseen test sets; classification methods on 9 datasets

Dataset	GMM	OCSVM	SVDD	Two SVDDs	SVM	RSVC
Fourclass	67.24 ±5.73	59.08 ±3.24	54.44 ±5.09	72.24 ±5.04	70.72 ±5.64	75.65 ±5.92
Liver disorders	40.86 ±5.63	43.48 ±4.88	47.68 ±4.4	50.25 ±5.15	52.03 ±3.03	54.12 ±5.53
Heart	46.33 ±4.26	57.49 ±4.83	46.41 ±4.24	61.08 ±3.02	71.51 ±4.47	72.12 ±4.36
Wine	33.43 ±5.03	42.09 ±6.99	21.41 ±2.35	46.46 ±5.84	75.66 ±4.86	76.99 ±4.69
Breast cancer	56.52 ±3.34	73.08 ±4.01	48.34 ±7.87	75.03 ±4.33	79.92 ±4.34	79.79 ±5.07
Diabetes	55.24 ±5.06	55.68 ±5.34	39.30 ±4.98	54.10 ±5.71	60.21 ±3.16	59.00 ±3.81
Australian	54.15 ±5.84	56.44 ±5.53	48.38 ±5.03	55.75 ±3.83	69.71 ±3.42	68.95 ±3.53
Inosphere	57.12 ±3.61	62.55 ±4.71	38.41 ±2.7	65.79 ±5.72	69.07 ±4.3	70.74 ±4.63
German numer	40.09 ±5.49	58.07 ±5.47	58.40 ±5.34	57.46 ±5.32	62.30 ±5.7	63.90 ±5.67

5 Conclusion

We have proposed the repulsive-SVDD classification to extend SVDD for binary classification problems. Two hyperspheres are trained in an optimisation problem to describe the distribution of two classes. Additional requirements are added to the optimisation problem to help with the discrimination task. First, the distance between two hypersphere centres is maximised to allow hyperspheres to expand. Second, margins between the hypersphere surfaces and data are maximised. The resulting method can create a decision boundary that takes information not only from distributions of the classes but also the boundary's margins. Experimental results on 9 datasets validate the good performance of the proposed method.

References

1. Tax, D.M.J., Duin, R.P.W.: Support vector data description. Machine Learning **54**(1), 45–66 (2004)
2. Schlkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C.: Estimating the support of a high-dimensional distribution. Neural Computation **13**(7), 1443–1471 (2001)
3. Wu, M., Ye, J.: A small sphere and large margin approach for novelty detection using training data with outliers. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(11), 2088–2092 (2009)
4. Le, T., Tran, D., Ma, W., Sharma, D.: An optimal sphere and two large margins approach for novelty detection. In: The 2010 International Joint Conference on Neural Networks (IJCNN), pp. 1–6. IEEE (2010)
5. Wang, J., Neskovic, P., Cooper, L.N.: Pattern classification via single spheres. In: Hoffmann, A., Motoda, H., Scheffer, T. (eds.) DS 2005. LNCS (LNAI), vol. 3735, pp. 241–252. Springer, Heidelberg (2005)
6. Hao, P.-Y., Chiang, J.-H., Lin, Y.-H.: A new maximal-margin spherical-structured multi-class support vector machine. Applied Intelligence **30**(2), 98–111 (2009)
7. Huang, G., Chen, H., Zhou, Z., Yin, F., Guo, K.: Two-class support vector data description. Pattern Recognition **44**(2), 320–329 (2011)
8. Schlkopf, B., Smola, A.J., Williamson, R.C., Bartlett, P.L.: New support vector algorithms. Neural Computation **12**(5), 1207–1245 (2000)
9. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) **2**(3), 27 (2011)
10. Bilmes, J.A., et al.: A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. International Computer Science Institute **4**(510), 126 (1998)